Rate Adaptation with Correlated Multi-Armed Bandits in 802.11 Systems

Yuzhou Tong*, Jiakun Fan*, Xuhong Cai*, and Yi Chen*†
*School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen
†Shenzhen Research Institute of Big Data
{yuzhoutong, jiakunfan, xuhongcai}@link.cuhk.edu.cn, yichen@cuhk.edu.cn

***Abstract*—Rate adaptation is a mechanism critical for maximizing the throughput of 802.11 systems. In this paper, inspired by the observation that the packet success rate under different data rates is not independent, we model the rate adaptation problem as a multi-armed bandit (MAB) problem with correlated arms. Two MAB-based rate adaptation algorithms are developed in which the correlation between data rates is exploited to accelerate the convergence of the algorithms. To verify the performance of our algorithms, we build up an indoor 802.11n test bed. The proposed rate adaptation algorithms are implemented and deployed on the test bed. Both simulation and test-bed experiments demonstrate the superiority of our algorithms in stationary and non-stationary radio environments.**

I. Introduction

Rate adaptation is an essential mechanism in 802.11 or Wi-Fi systems that adapts the physical data rate to the time-varying radio environment in order to maximize the expected throughput. In current 802.11 systems, the physical data rate depends on parameters including the Modulation and Coding Scheme (MCS), channel bandwidth, guard interval, and the number of spatial streams. As a result, the rate adaptation module is responsible for tuning these parameters simultaneously and selecting rates that yield the highest expected throughput from a large set of candidate rates.

Existing rate adaptation schemes can be broadly classified into two categories: one that relies on channel estimation and another on rate sampling. In the first category [1]–[5], the channel state information (CSI) such as Received Signal Strength Indicator (RSSI) or Signal-to-Noise Ratio (SNR) is assumed to be available. The CSI is used to estimate the probability of successful packet transmission at different rates, based on which a suitable rate is then selected. The major disadvantage of CSI-based rate adaptation is that system performance can be significantly reduced if the CSI measurements are inaccurate. In addition, most commercial Wi-Fi devices do not support CSI feedback at the transmitter.

In the absence of CSI, the second type of rate adaptation scheme estimates the performance of each rate directly by sampling. Different rates are sampled for packet transmission, and the acknowledgement (ACK) feedback is used to estimate the probability of success of the corresponding rate. The optimal rates are learned over time from the history of transmissions and outcomes. Due to its simplicity, this sampling-based rate adaptation is widely employed in current Wi-Fi systems, and popular algorithms include Minstrel [6] and its variant Minstrel HT [7], SampleRate [8], Auto Rate Fallback [9], etc. These algorithms primarily use well-engineered heuristics to sample rates. Examples of common rules include increasing rate upon success, decreasing rate upon failure, and assessing unused rates randomly using probe packets. While heuristic algorithms work well in simple scenarios, their efficiency has been questioned as Wi-Fi standards evolve and much more rates are available.

The key to sampling-based rate adaptation is the balance between exploration and exploitation. On the one hand, it needs to explore different rates to obtain more accurate information. On the other hand, it needs to exploit this information to transmit at optimal rates to gain throughput. As the number of available rates increases, it becomes more challenging to strike a balance between exploration and exploitation.

To overcome this challenge, researchers have recently considered introducing machine learning techniques like reinforcement learning [10], [11], neural network [12], and bayesian learning [13] to rate adaptation. Among these methods, the MAB algorithms such as Thompson Sampling (TS) [14]–[16] and Kullback-Leibler Upper-Confidence Bound (KL-UCB) [17] are especially popular due to their efficient sampling mechanism. However, these classical bandit algorithms exhibit two main limitations in the rate adaptation problem. First, the classical bandit algorithms assume that the rewards of different rates are independent, which is unlikely true. We know that the success probabilities of different rate are correlated. Second, the algorithms are usually designed for stationary environments, assuming each rate's reward follows a fixed distribution. However, in most cases, the wireless channel is non-stationary, and the reward distribution is time-varying. Regarding the first limitation, [17] uses a graph to model the correlations between rates in an 802.11n system. But the graph is constructed using observations and empirical results, and therefore is unable to be applicable to other radio environments or different Wi-Fi standards. Regarding the second limitation, [18] studied a contextual bandit model with CSI as a context.

In this paper, we propose a class of rate adaptation algorithms with correlated MAB. The main contributions of the paper are summarized as follows.

- We introduce a correlated MAB model for the rate adaptation problem. The correlation between different rates is measured as *pseudo-rewards*. We construct the *pseudo-rewards* information according to some prior knowledge

about the rate adaptation problem. Under the correlated MAB model, we develop two rate adaptation algorithms, C-KLUCB and C-TS, based on KL-UCB and TS, respectively. For non-stationary environments, we apply a sliding-window mechanism on C-KLUCB and C-TS to track the optimal rate online.
- We set up an indoor 802.11n test bed to evaluate the actual performance of different rate adaptation algorithms. The rate adaptation module is re-designed into two parts, one in the kernel space and the other in the user space. This separated design allows different rate adaptation algorithms to be efficiently implemented and deployed in the user space.
- We conduct both simulation and test-bed experiments in different channel environments to compare the performance of the proposed rate adaptation algorithms with some existing rate adaptation algorithms. The results show that our algorithms have an advantage in speed in identifying the appropriate rate.

II. System Model and Problem Formulation

We consider a single link consisting of a pair of transmitter and receiver implementing the IEEE 802.11 standard. Time is divided into slots which are indexed by $t = 1, 2, \ldots$. At each time slot t, let $h(t)$ denote the channel state and $r(t) \in \mathcal{R}$ denote the selected rate for packet transmission, where $\mathcal{R}$ denotes the set of available rates. Furthermore, let $p_r(t)$ denote the probability of successful transmission at rate r at time slot t. The objective of rate adaptation is to maximize the expected throughput by having the link transmit at the optimal rate in each time slot, that is,

$$r^*(t) = \arg\max_{r(t)\in\mathcal{R}} p_{r(t)}(t) \cdot r(t).$$

However, in practice, the probabilities $p_r(t)$ for all $r \in \mathcal{R}$ are unknown in advance, necessitating the use of online learning algorithms that learn the optimal rate over time from observed past transmission successes and failures.

We denote by $X_r(t)$ the outcome of the transmission at rate r at time slot t. Specifically, $X_r(t) = 1$ indicates a successful transmission and $X_r(t) = 0$ indicates a failed transmission. So $X_r(t)$ is a Bernoulli random variable with parameter $p_r(t)$, and we have $\mathbb{E}[X_r(t)] = p_r(t)$. At the end of each time slot, $X_r(t)$ is fed back to the transmitter to facilitate its decision on the rate for the next time slot. Let π denote an online learning policy that outputs a sequence of selected rates $r_\pi(1), r_\pi(2), \ldots, r_\pi(T)$ over a given time horizon T. Note that the choices $r_\pi(t)$ might be stochastic. The expected throughput under policy π can be found by

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} r_\pi(t) \cdot X_{r_\pi(t)}(t)\right] = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[r_\pi(t) \cdot p_{r_\pi(t)}(t)]$$

where the expectation on the left side of the equation is taken with respect to the random draw of both transmission outcome and the policy's selection of rate.

To quantify the performance of policy π, we consider the pseudo-regret, which is defined as the loss in expected throughput due to transmission at sub-optimal rates, i.e.,

$$\overline{Reg}(T) = \frac{1}{T}\sum_{t=1}^{T} r^*(t) \cdot p_{r^*(t)}(t) - \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[r_\pi(t) \cdot p_{r_\pi(t)}(t)]. \tag{1}$$

In a special case when the environment is stationary, and the channel state is time-invariant, i.e., $h(t) = h(t')$ for $t \neq t'$, we expect that the success probabilities of rates remain constant, i.e., $p_r(t) = p_r$ for $t = 1, \ldots, T$ and for all $r \in \mathcal{R}$. In such case, the optimal rate remains the same in all time slots, which is given by $r^* = \arg\max_{r\in\mathcal{R}} p_r \cdot r$. The goal of rate adaptation reduces to identify the best rate. The pseudo-regret can be written as

$$\begin{aligned}\overline{Reg}(T) &= r^* \cdot p_{r^*} - \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[r_\pi(t) \cdot p_{r_\pi(t)}] \\ &= \frac{1}{T}\sum_{r\in\mathcal{R}} (r^* - r)\mathbb{E}[n_r^\pi(T)],\end{aligned} \tag{2}$$

where $n_r^\pi(T)$ denotes the attempt number of rate r following policy π over time horizon T.

We now formulate the rate adaptation problem as finding a policy π that minimizes the pseudo-regret $\overline{Reg}(T)$, which is given by (1) in a non-stationary environment and (2) in a stationary environment. It is expected that in a real communication environment, the channel state is unlikely to change significantly in all time slots. Therefore, for a non-stationary environment, we assume that the channel state changes slowly; namely, the environment is piece-wise stationary, allowing us to design rate adaptation algorithms to efficiently track the optimal rate for transmission.

The challenge of the rate adaptation problem lies in balancing between staying with the rate that yielded the highest throughput in the past and exploring new rates that might yield higher throughput in the future. In the next section, we model the problem under the framework of the multi-armed bandit, which naturally addresses the fundamental trade-off between exploration and exploitation. In particular, some structure of the problem is exploited to reduce the amount of exploration and speed up the search for the optimal rate.

III. Rate adaptation with Correlated Bandit

We model the rate adaptation problem as a multi-armed bandit problem. Each candidate rate $r \in \mathcal{R}$ is treated as an arm, and the reward of the arm r at time slot t is $X_r(t)$ following a Bernoulli distribution with parameter $p_r(t)$. In the classical multi-armed bandit setting, the rewards associated with different arms are assumed to be independent, that is, the observed reward of pulling an arm r does not provide any information about the reward distribution of another arm r'. However, this is not the case in the rate adaptation problem, as the success probabilities of different rates are related. For instance, given a channel state $h(t)$, if a transmission at rate r is successful, then transmissions at rates smaller than r are

likely to be successful as well. We take advantage of this property and study the rate adaptation problem as the multi-armed bandit with correlated arms.

In the following, we will first describe the correlated multi-armed bandit model and introduce the concept of *pseudo-rewards*. Next, we will discuss the construction of *pseudo-rewards*, followed by rate adaptation algorithms. For ease of presentation, we focus on the stationary environment in Section III-A to C and discuss the non-stationary environment at the end in Section III-D.

A. Correlated multi-armed bandit model

In the multi-armed bandit problem, the mean rewards p_r of all arms, i.e., rates $r \in \mathcal{R}$ are unknown. In each time slot, the algorithm chooses an arm r and collects a random reward $X_r \in \{0, 1\}$ for this arm. Note that $\mathbb{E}[X_r] = p_r$. As mentioned earlier, the classical bandit model assumes that the mean rewards of different arms are independent. Therefore the collected reward X_r is only used to update the estimate of p_r but not the estimate of $p_{r'}$ for $r' \neq r$. In other words, the estimate of the mean reward of an arm is updated only when that arm is chosen. Such practice could be inefficient if the rewards across the arms are correlated, e.g., X_r can also be used to update the estimate of $p_{r'}$ when $X_{r'}$ and X_r are correlated.

Given two correlated arms r and r', according to the law of total expectation, we know that

$$\begin{aligned}\mathbb{E}[X_{r'}] &= \mathbb{E}[X_{r'}|X_r = 1] \cdot Pr\{X_r = 1\} \\ &\quad + \mathbb{E}[X_{r'}|X_r = 0] \cdot Pr\{X_r = 0\} \\ &= \mathbb{E}[X_{r'}|X_r = 1] \cdot \mathbb{E}[X_r] \\ &\quad + \mathbb{E}[X_{r'}|X_r = 0] \cdot (1 - \mathbb{E}[X_r]).\end{aligned} \tag{3}$$

So we can approximate $\mathbb{E}[X_{r'}]$ if we have some information of $\mathbb{E}[X_r]$, $\mathbb{E}[X_{r'}|X_r = 1]$ and $\mathbb{E}[X_{r'}|X_r = 0]$. The bandit algorithm is estimating $\mathbb{E}[X_r]$. But regarding $\mathbb{E}[X_{r'}|X_r = 1]$ and $\mathbb{E}[X_{r'}|X_r = 0]$, in reality, they are impossible to be acquired since two arms cannot be pulled at the same time. Still, we can extract some information about them based on domain knowledge. For instance, if a transmission at rate r is successful, then with high probability transmissions at rates smaller than r are also successful, which suggests $\mathbb{E}[X_{r'}|X_r = 1] \approx 1$ for $r' < r$. Following [19], we call these knowledge-based information as *pseudo-rewards*. We denote by $\gamma_{r'|r}(x)$ the pseudo-reward of arm r' conditional on $X_r = x$ for $x = 1$ or $x = 0$. In the next subsection, we will discuss some heuristics to construct the pseudo-rewards.

By (3), given some pseudo-rewards $\gamma_{r'|r}(1)$ and $\gamma_{r'|r}(0)$, an estimate of $\mathbb{E}[X_{r'}]$ with reference to arm r, denoted by $\tilde{p}_{r'|r}$, can be computed by

$$\tilde{p}_{r'|r} = \gamma_{r'|r}(1) \cdot \tilde{p}_r + \gamma_{r'|r}(0) \cdot (1 - \tilde{p}_r), \tag{4}$$

where $\tilde{p}_r$ is the empirical success probability of rate r learned from the collected outcomes of pulling arm r. By default, we set $\gamma_{r'|r'}(1) = 1$ and $\gamma_{r'|r'}(0) = 0$. As a result, $\tilde{p}_{r'|r'} = \tilde{p}_{r'}$.

We see that $\tilde{p}_{r'|r}$ will be updated once $\tilde{p}_r$ is updated, i.e., when arm r is pulled. Therefore in every time slot, there is always one of $\tilde{p}_{r'|r}$ for $r \in \mathcal{R}$ that will be updated. We use $\tilde{p}_{r'|r}$ for $r \in \mathcal{R}$ together as a reference to decide whether arm r' is worth exploring. The rationale behind the idea is that $\tilde{p}_{r'|r}$ is updated more frequently than $\tilde{p}_{r'}$, thus providing more fresh information about arm r'. Even though the fresh information can be very sketchy, it can help to reduce the amount of exploration. Details of the algorithm will be explained in Section III-C.

B. Construction of pseudo-reward

Since $\tilde{p}_{r'|r}$ is just a reference for exploration, its accuracy is not required, and so are the pseudo-rewards. We incorporate the domain knowledge about correlations between arms in the construction of pseudo-reward and make the following assumptions:

(1) If a transmission at rate r succeeds, transmissions at lower rates r' for $r' < r$ will also succeed.
(2) If a transmission at rate r succeeds, transmissions at higher rates r' for $r' > r$ may also succeed.
(3) If a transmission at rate r fails, transmissions at higher rates r' for $r' > r$ will also fail.
(4) If a transmission at rate r fails, transmissions at lower rates r' for $r' < r$ may still succeed.

Accordingly, for any $r, r' \in \mathcal{R}$, the pseudo-rewards are given by

$$\gamma_{r'|r}(1) = 1, \quad \gamma_{r'|r}(0) = \begin{cases} 0, & r' \geq r \\ 1, & r' < r. \end{cases}$$

Such choices of pseudo-rewards provide an optimistic estimate of $\tilde{p}_{r'|r}$, so as not to miss the exploration of certain potential rates. It needs to be mentioned that when all pseudo-rewards are set to 1, $\tilde{p}_{r'|r} = 1$, offering no extra information than $\tilde{p}_{r'}$, and thus the correlated bandit algorithm will perform the same as the standard bandit algorithm.

Finally, for notation simplicity, we use two matrices Γ_0 and Γ_1 of dimension $|\mathcal{R}| \times |\mathcal{R}|$ to describe the complete pseudo-rewards of all pairs of rates. The (i, j)-th entries in Γ_0 and Γ_1 correspond to $\gamma_{r_i|r_j}(0)$ and $\gamma_{r_i|r_j}(1)$, respectively.

C. Correlated bandit algorithms for rate adaptation

We now introduce the correlated bandit algorithms which borrow the idea from [19]. Compared to the standard bandit algorithms, the main difference of these algorithms is that they shrink the search space of optimal rate with the help of $\tilde{p}_{r'|r}$.

To be specific, in each time slot t, the algorithm maintains the following parameters:

- $\tilde{p}_r(t)$, $r \in \mathcal{R}$: the empirical success probability of rate r.
- $\tilde{p}_{r'|r}(t)$, $r, r' \in \mathcal{R}$: the estimate of mean reward of rate r' with reference to rate r.
- $n_r(t)$: the number of times rate r is chosen up to time t.
- $n_r^s(t)$: the number of times rate r is chosen and the transmission is successful.
- $\mathcal{S}_t = \{r \in \mathcal{R} : n_r(t) \geq \frac{t}{|\mathcal{R}|}\}$: the set of important rates.
- $\mathcal{A}_t \subseteq \mathcal{R}$: the set of candidate rates for exploration.

The empirically optimal rate is $\tilde{r}^*(t) = \arg\max_{r \in \mathcal{S}_t} r \cdot \tilde{p}_r(t)$. We define $\mathcal{A}_t = \{r \in \mathcal{R} : r = \tilde{r}^*(t) \text{ or } \min_{k \in \mathcal{S}_t} r \cdot \tilde{p}_{r|k}(t) \geq \tilde{r}^*(t) \cdot \tilde{p}_{\tilde{r}^*(t)}(t)\}$. Then KL-UCB and Thompson sampling can be used to sample a rate from $\mathcal{A}_t$ to execute.

In KL-UCB, the upper-confidence bound of rate r's success probability is computed by:

$$q_r(t) = \max\Big\{q \in [0,1] : n_k(t) \cdot d\Big(\frac{n_k^s(t)}{n_k(t)}, q\Big) \leq \log(t) + c \cdot \log\big(\log(t)\big)\Big\}, \quad (5)$$

where $c \geq 3$ is a positive constant, $d(x, y)$ denotes the KL divergence between two Bernoulli distributions with means x, y and is defined as $d(x, y) = x \cdot \log(\frac{x}{y}) + (1-x) \cdot \log(\frac{1-x}{1-y})$. The correlated KL-UCB algorithm is summarized in Algorithm 1.

The correlated TS algorithm is similar except that it models the success probability $\tilde{p}_r(t)$ by a Beta distribution. Details of the algorithm are shown in Algorithm 2.

In summary, the key innovation of the two rate adaptation algorithms is that they exploit the correlation structure of rates to improve the efficiency of exploration.

Algorithm 1 Correlated KL-UCB

Inputs: Pseudo-rewards Γ_0 and Γ_1
Initialize: For $t = 1, \ldots, |\mathcal{R}|$, play each rate $r \in \mathcal{R}$ once.
$\tilde{p}_r(|\mathcal{R}|) \leftarrow n_r^s(|\mathcal{R}|)$ and $\mathcal{S}_{|\mathcal{R}|} \leftarrow \mathcal{R}$
Compute $\tilde{p}_{r|k}(|\mathcal{R}|)$ for $r, k \in \mathcal{R}$ using (4)
$\tilde{r}^*(|\mathcal{R}|) \leftarrow \arg\max_{r \in \mathcal{S}_{|\mathcal{R}|}} r \cdot \tilde{p}_r(|\mathcal{R}|)$
$\mathcal{A}_{|\mathcal{R}|} \leftarrow \{r \in \mathcal{R} : \min_{k \in \mathcal{S}_{|\mathcal{R}|}} r \cdot \tilde{p}_{r|k}(|\mathcal{R}|) \geq \tilde{r}^*(|\mathcal{R}|) \cdot \tilde{p}_{\tilde{r}^*(|\mathcal{R}|)}(|\mathcal{R}|)\} \cup \{\tilde{r}^*(|\mathcal{R}|)\}$.
for $t = |\mathcal{R}| + 1, ..., T$ **do**
  for each rate $r \in \mathcal{A}_{t-1}$ **do**
    Compute $q_r(t-1)$ using (5)
  end for
  $r^* \leftarrow \arg\max_{k \in \mathcal{A}_{t-1}} k \cdot q_k(t-1)$
  Select rate r^* and observe X_{r^*}
  for each rate $r \in \mathcal{R}$ **do**
    if $r = r^*$ **then**
      $n_r(t) \leftarrow n_r(t-1) + 1$
      $n_r^s(t) \leftarrow n_r^s(t-1) + X_{r^*}$
    else
      $n_r(t) \leftarrow n_r(t-1)$
      $n_r^s(t) \leftarrow n_r^s(t-1)$
    end if
    $\tilde{p}_r(t) \leftarrow \frac{n_r^s(t)}{n_r(t)}$
    Update $\tilde{p}_{r|k}(t)$ using (4)
  end for
  $\mathcal{S}_t \leftarrow \big\{r \in \mathcal{R} : n_r(t) \geq \frac{t}{|\mathcal{R}|}\big\}$
  $\tilde{r}^*(t) \leftarrow \arg\max_{r \in \mathcal{S}_t} r \cdot \tilde{p}_r(t)$
  $\mathcal{A}_t \leftarrow \{r \in \mathcal{R} : r = \tilde{r}^*(t) \text{ or } \min_{k \in \mathcal{S}_t} r \cdot \tilde{p}_{r|k}(t) \geq \tilde{r}^*(t) \cdot \tilde{p}_{\tilde{r}^*(t)}(t)\}$
end for

D. Non-stationary Environment.

For the non-stationary environment, we consider it to be piece-wise stationary. The success probability of each rate

Algorithm 2 Correlated TS

Inputs: Pseudo-rewards Γ_0 and Γ_1
Initialize: $n_r^s(0) = 1, n_r(0) = 2$ for $r \in \mathcal{R}$
For $t = 1, \ldots, |\mathcal{R}|$, play each rate $r \in \mathcal{R}$ once.
$n_r(|\mathcal{R}|) \leftarrow n_r(0) + 1$ and $n_r^s(|\mathcal{R}|) \leftarrow n_r^s(0) + X_r$
Sample $\tilde{p}_r(|\mathcal{R}|) \sim Beta\big(n_r^s(|\mathcal{R}|), n_r(|\mathcal{R}|) - n_r^s(|\mathcal{R}|)\big)$
Same as steps 4-6 in Algorithm 1
for $t = |\mathcal{R}| + 1, ..., T$ **do**
  $r^* \leftarrow \arg\max_{k \in \mathcal{A}_{t-1}} k \cdot \tilde{p}_k(t-1)$
  Same as steps 12-20 in Algorithm 1
  Sample $\tilde{p}_r(t) \sim Beta\big(n_r^s(t), n_r(t) - n_r^s(t)\big)$
  Same as steps 22-26 in Algorithm 1
end for

$p_r(t)$ evolves slowly over time. We employ the Sliding-Window (SW) mechanism to track the optimal rate. In this mechanism, the rate adaptation algorithms only count the statistics of each rate within a time window of size w. In particular, the number of times rate k is selected is computed as follows

$$n_k(t) = \sum_{\tau=t-w}^{t} \mathbb{1}(r(\tau) = k)$$
$$n_k^s(t) = \sum_{\tau=t-w}^{t} X_k(\tau)\mathbb{1}(r(\tau) = k),$$

where $\mathbb{1}(\cdot)$ is an indicator function. Algorithm 1 and 2 still work by replacing the computation of $n_k(t)$ and $n_k^s(t)$ as described above. We choose the window size w carefully. On the one hand, w needs to be large enough to estimate each rate's success probability accurately. On the other hand, w cannot be too large such that it fails to track the changes in the success probabilities of the rates. In addition, we would like to emphasize that the pseudo-rewards constructed in Section III-B are generic and applicable to non-stationary environments. With the help of pseudo-rewards, the correlated bandit algorithm is able to adjust its estimates of the rates more quickly when the channel changes. In the next section, we will verify this experimentally.

IV. Simulation Results

In this section, we evaluate four rate adaptation algorithms: TS, KL-UCB, correlated TS (C-TS), and correlated KL-UCB (C-KLUCB) by simulation. The simulation setting refers to 802.11n systems with eight available rates, i.e., $\mathcal{R} = \{6, 9, 12, 18, 24, 36, 48, 54\}$ Mbps. Like [8], [17], we consider three typical scenarios under stationary environments: *steep*, *gradual* and *lossy*. In steep scenarios, the success probability is either very high or very low. In gradual cases, the success probability gradually decreases as the rate increases, and the success probability of the optimal rate is higher than 0.5. While for the lossy cases, the success probability of the optimal rate is lower than 0.5. The success probabilities of the

eight rates under three scenarios are given below (the optimal rate is marked in bold).

$$\mathcal{P}_{steep} = \{.99, .98, .96, .93, \mathbf{.90}, .10, .06, .04\}$$
$$\mathcal{P}_{gradual} = \{.95, .90, .80, \mathbf{.65}, .45, .25, .15, .10\}$$
$$\mathcal{P}_{lossy} = \{.90, .80, .70, .55, .45, \mathbf{.35}, .20, .10\}$$

The cumulative regret of each algorithm is shown in Fig. 1. We have the following observations. First, in all scenarios, the regrets of the four algorithms grow sub-linearly, meaning that all algorithms can identify the optimal rate. This observation is consistent with the regret bound analysis of KL-UCB [20] and TS [21]. Second, C-TS and C-KLUCB have lower regrets than TS and KL-UCB, which demonstrates the superiority of the correlated bandit model.

V. Test-Bed Experiment

A. Test-bed implementation

We set up an indoor 802.11n test bed with an Access Point (AP) and a client. The client is implemented on *Dongtintech AIMB-B2205A* which supports *Ath9k* driver that by default uses the *Minstrel* algorithm for rate adaptation. The client runs Ubuntu 22.04 with a 5.15 kernel. The AP is a *Raspberry Pi 4B* running the Linux-based *OpenWrt* backfire 10.03 operating system with 5.4.132 Linux kernel. The AP uses the 802.11n Wi-Fi protocol to connect to the client. *Edimax EW-7833UAC* Wi-Fi adaptor is used which supports two groups of rates (one uses 400ns GI for the 8 MCS, and the other uses 800ns GI).

We deploy the four bandit-based rate adaptation algorithms on the client machine. Since some basic operations, such as floating point calculation and logarithms computation, are limited in the kernel space, we re-design the rate adaptation module. It consists of two parts: user space and kernel space. The bandit algorithms are implemented and deployed in the user space using *Python*. The user space is responsible for reading the debugfs file *rc_stats* offered by the 802.11 rate adaptation module and getting the success and attempt numbers of each rate. With these statistics, the bandit algorithms in the user space can be executed to select the rate for the subsequent transmission. The kernel space is a modified 802.11 rate adaptation module to receive and pass the selected rate to the driver. Fig. 2 shows the architecture. Our design significantly enhances the efficiency of test-bed implementation since we do not need to implement all the basic operations from scratching or compiling the kernel module for debugging. Meanwhile, this implementation also allows us to deploy and test different rate adaptation algorithms conveniently.

To create a non-stationary environment, we change the location of the AP and add some obstacles between the AP and the client during the test. We use the popular network measurement tool *iperf3* to perform the test with UDP traffic. The UDP sessions have a packet length of 8KB.

B. Experiment results

We first test the performance of the five rate adaptation algorithms: TS, KL-UCB, C-TS, C-KLUCB, and Minstrel in three stationary scenarios which are described in Table I. For each scenario, the test lasts for 60s. Fig. 3 shows the average throughput of each algorithm. It can be seen that the correlated rate adaptation algorithms outperform their classical versions in $E2$ and $E3$, while all algorithms achieve similar performance in $E1$.

TABLE I
Test-bed experiment scenarios

	Scenario	Frequency	Channel width
E1	The client is close to the AP	2.4 GHz	20 MHz
E2	The client is far from the AP	2.4 GHz	20 MHz
E3	Far from AP, With obstacle	2.4 GHz	20 MHz

Next, we test different rate adaptation algorithms in a non-stationary environment. The sliding window technique is used. To create such an environment, each scenario in Table I keeps for about the 60s and gradually changes to the next scenario within 10s ($E1 \rightarrow E2 \rightarrow E3$). Fig. 4 shows the instantaneous throughput of each algorithm. We observe that all the bandit rate adaptation algorithms are superior to Minstrel. Moreover, SW-C-TS and SW-C-KLUCB can track the variation of the radio environment more rapidly and achieve more stable throughput. Especially in the third scenario, SW-C-KLUCB and SW-C-TS can find the optimal rate faster than SW-KLUCB and SW-TS, demonstrating the gain by utilizing the correlation between different rates.

VI. Conclusion

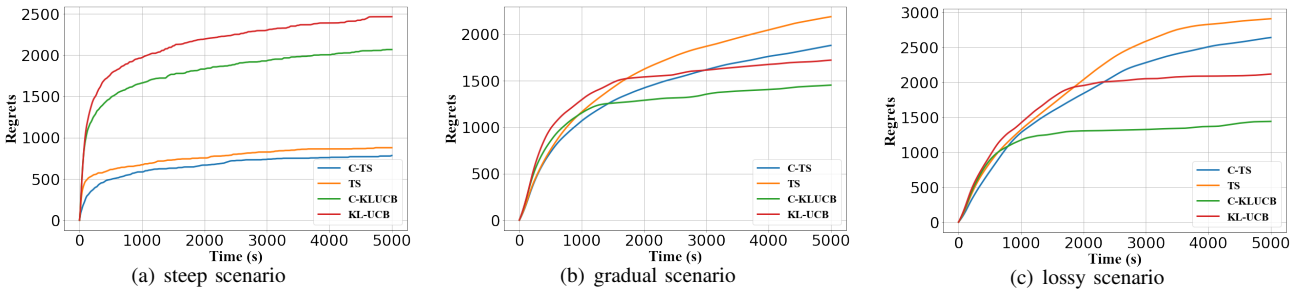
This paper presents a bandit-based rate adaptation model with correlated rates. The fact that the success probabilities of different rates are correlated is taken advantage in the form of *pseudo-rewards*. We consider four general rules for constructing pseudo-rewards based on domain knowledge so they are applicable to both stationary and non-stationary environments. These pseudo-rewards help to identify the optimal rates with less amount of exploration and thus can improve the sampling efficiency. Based on the correlated bandit model, two rate adaptation algorithms called C-KLUCB and C-TS are introduced. Simulation and test-bed experiments are both carried out to evaluate their performance. The results demonstrate their effectiveness and show that C-KLUCB and C-TS can provide significant superiority over benchmark algorithms, e.g., Minstrel in both stationary and non-stationary environments.

References

[1] J. Camp and E. Knightly, "Modulation Rate Adaptation in Urban and Vehicular Environments: Cross-Layer Implementation and Experimental Evaluation," *IEEE/ACM Transactions on Networking*, vol. 18, no. 6, pp. 1949–1962, Dec. 2010.

[2] L. Deek, E. Garcia-Villegas, E. Belding, S.-J. Lee, and K. Almeroth, "Joint rate and channel width adaptation for 802.11 MIMO wireless networks," in *2013 IEEE International Conference on Sensing, Communications and Networking (SECON)*, Jun. 2013, pp. 167–175.

[3] I. Pefkianakis, Y. Hu, S. H. Wong, H. Yang, and S. Lu, "MIMO rate adaptation in 802.11n wireless networks," in *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '10. New York, NY, USA: Association for Computing Machinery, Sep. 2010, pp. 257–268.



(a) steep scenario (b) gradual scenario (c) lossy scenario

Fig. 1. Regrets for TS, KL-UCB, C-TS, and C-KLUCB under three different stationary scenarios: (a) steep, (b) gradual, and (c) lossy scenarios.

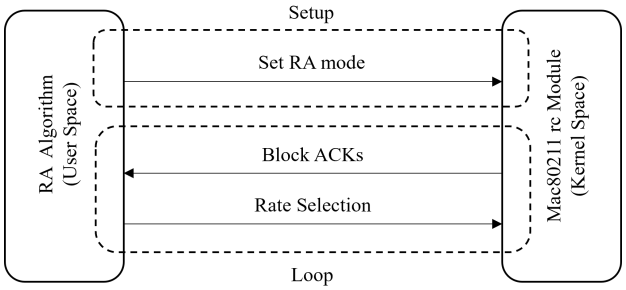


Fig. 2. Bandit rate adaptation architecture

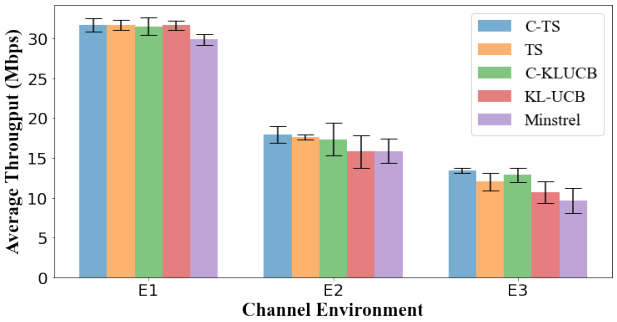


Fig. 3. Average throughput of each rate in different channel environments

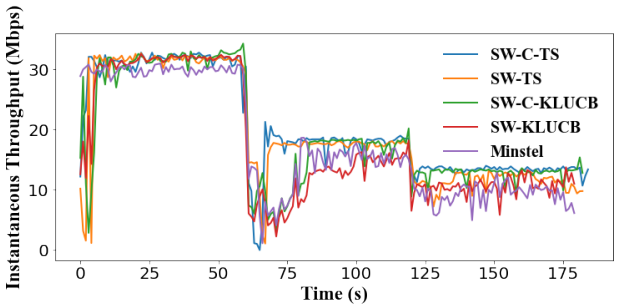


Fig. 4. Instantaneous throughput of the various rate adaptation algorithms in 802.11n systems in the non-stationary environment

[4] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," *ACM SIGCOMM computer communication review*, vol. 40, no. 4, pp. 159–170, 2010.

[5] J. Zhang, K. Tan, J. Zhao, H. Wu, and Y. Zhang, "A Practical SNR-Guided Rate Adaptation," in *IEEE INFOCOM 2008 - The 27th Conference on Computer Communications*, Apr. 2008, pp. 2083–2091.

[6] F. Fietkau, "The minstrel rate control algorithm for mac802.11," 2020.

[7] ——, "Minstrel_ht: new rate control module for 802.11n," 2010.

[8] J. C. J. C. Bicket, "Bit-rate selection in wireless networks," Thesis, Massachusetts Institute of Technology, 2005.

[9] A. Kamerman and L. Monteban, "WaveLAN®-II: A high-performance wireless LAN for the unlicensed band," *Bell Labs Technical Journal*, vol. 2, no. 3, pp. 118–133, 1997.

[10] S.-C. Chen, C.-Y. Li, and C.-H. Chiu, "An Experience Driven Design for IEEE 802.11ac Rate Adaptation based on Reinforcement Learning," in *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, May 2021, pp. 1–10.

[11] V. Saxena, H. Tullberg, and J. Jaldén, "Reinforcement Learning for Efficient and Tuning-Free Link Adaptation," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 768–780, Feb. 2022.

[12] S. Khastoo, T. Brecht, and A. Abedi, "NeuRA: Using Neural Networks to Improve WiFi Rate Adaptation," in *Proceedings of the 23rd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, ser. MSWiM '20. New York, NY, USA: Association for Computing Machinery, Nov. 2020, pp. 161–170.

[13] V. Saxena and J. Jaldén, "Bayesian Link Adaptation under a BLER Target," in *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, May 2020, pp. 1–5.

[14] H. Qi, Z. Hu, X. Wen, and Z. Lu, "Rate Adaptation With Thompson Sampling in 802.11ac WLAN," *IEEE Communications Letters*, vol. 23, no. 10, pp. 1888–1892, Oct. 2019.

[15] H. Gupta, A. Eryilmaz, and R. Srikant, "Link Rate Selection using Constrained Thompson Sampling," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, Apr. 2019, pp. 739–747.

[16] ——, "Low-Complexity, Low-Regret Link Rate Selection in Rapidly-Varying Wireless Channels," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, Apr. 2018, pp. 540–548.

[17] R. Combes, J. Ok, A. Proutiere, D. Yun, and Y. Yi, "Optimal Rate Sampling in 802.11 Systems: Theory, Design, and Implementation," *IEEE Transactions on Mobile Computing*, vol. 18, no. 5, pp. 1145–1158, May 2019.

[18] A. Sen and K. M. Sivalingam, "Rate adaptation techniques using contextual bandit approach for mobile wireless lan users," in *2020 IEEE 45th Conference on Local Computer Networks (LCN)*. Los Alamitos, CA, USA: IEEE Computer Society, nov 2020, pp. 469–472. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/LCN48667.2020.9314810

[19] S. Gupta, S. Chaudhari, G. Joshi, and O. Yağan, "Multi-Armed Bandits With Correlated Arms," *IEEE Transactions on Information Theory*, vol. 67, no. 10, pp. 6711–6732, Oct. 2021.

[20] A. Garivier and O. Cappé, "The KL-UCB Algorithm for Bounded Stochastic Bandits and Beyond," in *Annual Conference Computational Learning Theory*, Feb. 2011.

[21] S. Agrawal and N. Goyal, "Analysis of Thompson Sampling for the Multi-armed Bandit Problem," in *Proceedings of the 25th Annual Conference on Learning Theory*. JMLR Workshop and Conference Proceedings, Jun. 2012, pp. 39.1–39.26.